OUT-OF-SAMPLE FORECAST ERRORS IN MISSPECIFIED PERTURBED LONG MEMORY PROCESSES

Francesc Marmol and
Miguel A. Arranz

98-49

WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 98-49
Statistics and Econometrics Series 21
July 1998

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34-91) 624-9849

OUT-OF-SAMPLE FORECAST ERRORS IN MISSPECIFIED PERTURBED LONG MEMORY PROCESSES

Marmol Francesc and Arranz, Miguel A. *

Abstract

The correlogram is not a useful diagnosis tool in the presence of long-memory or long range depedent time series. The aim of this paper is to illustrate this claim by examining the relative increase in mean square forecast error from fitting a weakly stationary process to the series of interest hen in fact the true model is a so-called perturbed long-memory process recently introduced by Granger and Marmol (1997). This model has the property of being unidentifiable from a white noise process on the basis of the correlogram and the usual rule-of thumbs in the Box-Jenkins methodology. We prove that this kind of misspecification can lead to serious errors in terms of forecasting.

Key Words

Perturbed long-memory; correlogram; forecast error.

*Department of Statistics y Econometrics, Universidad Carlos III de Madrid, e-mail: fmarmol@est-econ.uc3m.es ; Department of Economics, Universidad de Alicante. The authors wold like to thank J. Gonzalo for his many suggestions and comments.

1 Introduction

There are at least three different reasons not to consider the correlogram as a useful diagnostic tool in the presence of long memory or long range dependent time series. First, it is widely known that long memory processes have autocorrelations that decay at a slow rate, proportional to $k^{-\alpha}$, $\alpha \in (0,1)$ as lag k goes to infinity. Thus, the correlogram should exhibit this slow decay. However, for values of α close to one, it is very difficult to distinguish between long and short memory processes, i.e. between processes with autocorrelations decaying at a hyperbolic rate and processes with autocorrelations dying out at the faster exponential rate. Second, since long memory is an asymptotic notion, we should look at the correlogram at high lags which, in turn, cannot be estimated in a reliable way. Third, the above definition of long memory only implies that the autocorrelations decay so slowly that are not in fact absolutely summable. However, the absolute values of the individual autocorrelations can be arbitrarily small.

All these difficulties are well reported (see Hampel et al. 1986, Chapter 8.1c, and Beran 1994, Chapter 4.3) and are likely to lead the practitioner to severely unsuitable conclusions. In this sense, the question of the effects of misspecification errors with long memory processes has been recently discussed, among others, by Yajima (1993) and Ray (1993) for the case where we fit a finite-order autoregressive model to a stationary *ARFIMA* process. On the other hand, Hassler (1994) discussed the misspecification of a long memory process in seasonal time series whereas Crato and Taylor (1996) studied the case in which a stationary *ARFIMA* model is misspecified as a nonstationary *ARIMA* process.

Granger and Marmol (1997) have recently added a new reason against the use of the correlogram as an heuristic device to detecting long run dependence. They consider the sum of a long memory process plus an independent white noise component and prove that this new process, called *perturbed long memory* (*PLM*) process, has arbitrarily small autocorrelations, depending on the underlying signal–to–noise ratio. In particular, they show that there might

caveat applies.

exist a signal–to–noise ratio for which the corresponding *PLM* process will have its autocorrelations inside the $\pm\frac{2}{\sqrt{T}}$ bands, with T denoting the number of observations. Thus, such a particular member of the *PLM* class of processes will have a correlogram indistinguishable from that of a white noise process (at a 5% level of significance) on the basis of the widely applied Box–Jenkins methodology, *uniformly on* $\alpha \in (0,1)$ and *for all* k.

Yet, it is a well known fact that the use of erroneous models has a great impact on, for instance, the accuracy of forecasts. This problem is particularly relevant when the true model is long range dependent, given that, as pointed out by Beran (1994), long–range correlations are relevant for statistical purposes, even in the case of small individual correlations, since the relevant feature is that the sum of correlations is large.

Consequently, the aim of this paper is to investigate the increase in forecast error due to the erroneous identification of a *PLM* model as a white noise process. We consider the particular case where the *PLM* model is a stationary *ARFIMA* process in order to provide closed analytical expressions.

The paper is organized as follows. In Section 2 we introduce the model of interest, while Section 3 provides the expressions of the increase in mean square forecast error due to the erroneous identification of a *PLM* as a white noise process. On the other hand, it is well known that we can use the $\pm\frac{2}{\sqrt{T}}$ bounds to roughly check the significance of the sample autocorrelations, and that we should look for regularities even if the sample autocorrelations are inside the $\pm\frac{2}{\sqrt{T}}$ limits. Hence, in Section 4 we study the increase in mean square error in the case that we fit an $AR(1)$ process. Finally, Section 5 concludes. Proofs of the different propositions are collected in Appendix A.

2 The Model and Implications

An *ARFIMA* (p, d, q) process $\{y_t\}$ is generated by

$$\phi_p(B)(1-B)^d y_t = \theta_q(B)u_t \tag{1}$$

where u_t is an independent and identically distributed white noise process with zero mean and finite variance σ_u^2. The *ARFIMA* model (Granger and Joyeux, 1980, Hosking, 1981) generalizes the well–known *ARIMA* model by allowing for non integer differencing powers. It can be proved that if the roots of the autoregressive and moving average polynomials lie outside the unit circle and $d < \frac{1}{2}$, y_t is stationary. When $d \in (0, \frac{1}{2})$ the Wold decomposition and autocorrelation coefficients of y_t will exhibit the slow hyperbolic rate of decay characterizing the long memory process. In particular, it can be shown that for high lags, $\rho_y(k) \sim ck^{2d-1}$, $c > 0$, so that the *ARFIMA* process (1) is a particular member of the long memory processes with $\alpha = 1-2d$. On the other hand, when $d = 0$, (1) becomes the standard $ARMA(p,q)$ model, a stationary process with autocorrelations decaying at an exponential rate. Throughout the paper, we shall only consider the case $d \in (0, \frac{1}{2})$.

Assume now that the basic *ARFIMA* process, y_t is observed with some sampling error

$$x_t = y_t + \epsilon_t \tag{2}$$

where along the paper ϵ_t is assumed to be an independent and identically distributed white noise process with zero mean and finite variance σ_ϵ^2, stochastically independent of u_t. Expression (2) is the model of interest considered by Granger and Marmol (1997). They show that the autocorrelation function of x_t is

$$\rho_x(k) = \frac{1}{1+\zeta}\rho_y(k) \qquad k = 1, 2, \dots , \tag{3}$$

where $\zeta = \sigma_\epsilon^2/\mathrm{var}(y_t)$, would be the corresponding signal–to–noise ratio. Therefore, $0 < \rho_x(k) < \rho_y(k)$, $k = 1, 2, \dots$, where the first inequality follows from the long memory nature of y_t. Furthermore, it also follows from (3) in a straightforward manner that

$$\rho_x(k) \sim \frac{c}{1+\zeta} k^{2d-1} \equiv c^\star k^{2d-1} \qquad\qquad k \to \infty, \quad c > c^\star, \tag{4}$$

so that x_t turns out to be a long memory process with greater variability than the signal y_t. Granger and Marmol (1997) called x_t in expression (2) a *perturbed long memory (PLM)* process, and the constant $c^\star$ the *size* of the process. It is also clear from (3) that there may exist a sufficiently large signal–to–noise ratio, $\tilde{\zeta}$, for which the autocorrelations of the *PLM* process, say $\tilde{x}_t$, will be arbitrarily small, in particular smaller than $\frac{2}{\sqrt{T}}$, for all k, $d \in (0, \frac{1}{2})$ and c, i.e. even in the case where the signal y_t is a long memory process with high persistence and large individual autocorrelations. It is thus obvious that $\tilde{x}_t$ will have near–observational problems in the sense of distinguishing between autocorrelations with hyperbolic decay from autocorrelations with exponential decay on the basis of the shape of the correlogram.

On the other hand, in empirical analysis we deal with sample autocorrelations rather than with the theoretical ones. In this sense, under some regularity conditions, Granger and Marmol (1997) prove that

$$E\Big(\hat{\rho}_x(k) - \rho_x(k)\Big) \sim \frac{-c^\star}{d(1+2d)}\Big(1 - \rho_x(k)\Big)T^{2d-1}, \tag{5}$$

so that the sample autocorrelations of the *PLM* process x_t actually underestimate their theoretical counterparts, with the negative bias decaying at a slower rate as the sample size T increases. Hence, the risk of misspecification increases when looking at the empirical correlogram.

To conclude this section, it is worth mentioning that the *PLM* model considered by Granger and Marmol belongs to a wider family of processes sharing the same kind of behavior. First, it is not difficult to extend the model to allow for a general short memory process as the perturbed factor instead of just white noise. Second, and just for illustration purposes, consider the so–called long–memory stochastic volatility process developed by Harvey (1993) and Breidt et al. (1998). The model is

$$x_t = \lambda_t \sigma_t, \quad \sigma_t^2 = \sigma^2 \exp(h_t), \quad (1-B)^d h_t = \eta_t, \quad \lambda_t \sim N(0,1), \quad \eta_t \sim N(0, \sigma_t^2).$$

After some manipulation, this model can be written as $\log x_t^2 = h_t + noise$, i.e. as a fractional white noise process plus some other noise. Thus, the results obtained in the next sections of the paper also apply to the aforementioned more general set–ups.

3 Misspecified *PLM* Processes

In this section we shall be concerned with the possible increase in forecast error due to the erroneous identification of the *PLM* model $\tilde{x}_t$ as a white noise process.

Assume that we have observations $x_1, \ldots, x_{T-1}, x_T$, of the true *PLM* process, x_t, and we would like to forecast x_{T+h}. Let $\hat{x}_{T+h|T}$ denote the minimum mean square error (mmse) linear predictor of x_{T+h} based on the values of x_j, $j \leq T$, i.e. $\hat{x}_{T+h|T} = E(x_{T+h}|x_T, x_{T-1}, \ldots)$, and let $e_{T+h|T} = x_{T+h} - \hat{x}_{T+h|T}$ be the corresponding forecast error. Analogously, we shall denote by $\hat{x}^\star_{T+h|T}$ and $e^\star_{T+h|T} = x_{T+h} - \hat{x}^\star_{T+h|T}$ the predictor and the forecast error for the misspecified white noise model, respectively. Herein we are interested in the relative increase in mean square forecast error,

$$\frac{\operatorname{var}\left(e^\star_{T+h|T}\right) - \operatorname{var}\left(e_{T+h|T}\right)}{\operatorname{var}\left(e_{T+h|T}\right)}, \tag{6}$$

especially in the case when the *PLM* has all its autocorrelations below $\frac{2}{\sqrt{T}}$, i.e., when $x_t = \tilde{x}_t$. We have the following result.

Proposition 1. *Let x_t be a* PLM *process with the signal y_t generated according to expression* (1), $d \in (0, \frac{1}{2})$. *Suppose that we take x_t as a white noise process. Then, for any ζ, the relative increase (*RI*) in mean square forecast error is given by*

$$RI(h) = \frac{var(y_t) - \sigma_u^2 \sum_{j=0}^{h-1} \psi_j^2}{\sigma_u^2 \sum_{j=0}^{h-1} \psi_j^2 + \sigma_\epsilon^2} = \frac{1 - \dfrac{\sigma_u^2}{var(y_t)} \sum_{j=0}^{h-1} \psi_j^2}{\dfrac{\sigma_u^2}{var(y_t)} \sum_{j=0}^{h-1} \psi_j^2 + \zeta}, \tag{7}$$

where ψ_j, $j = 0, 1, \ldots$ are the coefficients of the moving average representation of y_t.

From the general expression given in (6), we can obtain closer analytic expressions for the relative increase in mean square forecast error by considering two particular but very important cases of the $ARFIMA(p,d,q)$ processes, namely the $ARFIMA(0,d,0)$ and $ARFIMA(1,d,0)$ models, referred in the text as Case A and Case B, respectively.

Proposition 2. *Under the same assumptions as in Proposition 1, in Case A, with* $(1-B)^d y_t = u_t$,

$$RI_A(h) = \frac{\dfrac{\Gamma(1-2d)}{\Gamma^2(1-d)} - \sum_{j=0}^{h-1}\left(\dfrac{\Gamma(j+d)}{\Gamma(d)\Gamma(1+j)}\right)^2}{\sum_{j=0}^{h-1}\left(\dfrac{\Gamma(j+d)}{\Gamma(d)\Gamma(1+j)}\right)^2 + \beta}. \tag{8}$$

In Case B, with $(1-\phi B)(1-B)^d y_t = u_t$, $|\phi| < 1$, $\phi \neq 0$,

$$RI_B(h) = \frac{\dfrac{\Gamma(1-2d)}{\Gamma^2(1-d)}\dfrac{F(1,1+d,1-d;\phi)}{1+\phi} - \sum_{j=0}^{h-1}\left[\dfrac{\Gamma(j+d)}{\Gamma(d)\Gamma(1+j)}F(1,-j,1-d-j;\phi)\right]^2}{\sum_{j=0}^{h-1}\left[\dfrac{\Gamma(j+d)}{\Gamma(d)\Gamma(1+j)}F(1,-j,1-d-j;\phi)\right]^2 + \beta}, \tag{9}$$

where $\beta = \dfrac{\sigma_\epsilon^2}{\sigma_u^2}$, $\Gamma(z)$ *denotes the Gamma function and* $F(a,b,c;z)$ *is the Hypergeometric function.*

In order to illustrate the results obtained in the above propositions, we provide some visual evidence in Figures 1 and 2. From expressions (8) and (9) it is clear that the relative increase in the mean square forecast error depends positively on d and negatively on the ratio $\beta = \frac{\sigma_\epsilon^2}{\sigma_u^2}$. Herein we present two extreme cases, namely, $\{d = 0.4, \beta = 1.0\}$, and $\{d = 0.1, \beta = 5.0\}$, with $\phi = 0.4$ in Case B. A further complete set of experiments is available upon request. With the first (second) one we shall obtain large (small) relative increases in relative mean square forecast errors. On the other hand, it is worth mentioning that in our simulations, and for medium sized samples, the selected ratios $\beta = 1$, and $\beta = 5$, usually suffice to have all the autocorrelations of the corresponding PLM process within the $\pm\frac{2}{\sqrt{T}}$ bands, with the autocorrelations being closer to the bands in the second case.

Figure 1 about here

Figure 2 about here

Figures 1 and 2 both display very large relative increase in mean square forecast error in the $\{d = 0.4, \beta = 1.0\}$ case and for many periods ahead. For instance, in Case A we obtain that this relative increase is about 50% for short–range forecasts and still above 20% for 25 periods ahead. The situation is even worse for Case B. The combination of large values of the long–memory parameter d, low values of β and medium values of the autoregressive parameter ϕ lead to a relative increase of in the mean square forecast error close to 160% in the short–run and about 40% 25 periods ahead.

As it could be expected, the situation is not so serious in the second case, $\{d = 0.1, \beta = 5.0\}$, and in fact the relative increase in the mean square forecast errors is negligible in the long–run. However, even in this case, when the data generating process clearly resembles a white noise process, for medium values of the autoregressive parameter ϕ, the relative increase in the mean square forecast error is about 5% in the short–run.

4 Looking for Stationary Structures

It is widely accepted that we can use the $\pm\frac{2}{\sqrt{T}}$ bounds as a rough indicator of the significance of the sample autocorrelations, and that we should look for regularities even if the sample autocorrelations are inside those bounds. In this section we shall assume that the true data generating process is a *PLM* process with all of its autocorrelations within the $\pm\frac{2}{\sqrt{T}}$ bound but that, instead of a white noise process, the proposed model is now an autoregressive process. In the case that we fit an $AR(1)$, Propositions 1 and 2 are generalized as follows.

Proposition 3. *Let x_t be a* PLM *process with the signal y_t generated according to expression* (1), *$d \in (0, \frac{1}{2})$. Suppose that we model x_t as an $AR(1)$ process with coefficient ρ. Then, for any ζ, the relative increase in mean square forecast error is given by*

$$RI(h) = \frac{\sum_{j=h}^{\infty}(\psi_j - \rho^h \psi_{j-h})^2 + \rho^{2h}\beta}{\sum_{j=0}^{h-1} \psi_j^2 + \beta} \tag{10}$$

Proposition 4. *Under the same assumptions as in Proposition 3, in Case A, with* $(1-B)^d y_t = u_t$,

$$RI_A(h) = \frac{\dfrac{(1+\rho_A^{2h})\Gamma(1-2d)}{\Gamma^2(1-d)} - \dfrac{2\rho_A^h(-1)^h\Gamma(1-2d)}{\Gamma(1+h-d)\Gamma(1-h-d)} + 2\rho_A^{2h}\beta - \sum_{j=0}^{h-1}\left(\dfrac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}\right)^2}{\sum_{j=0}^{h-1}\left(\dfrac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}\right)^2 + \beta}. \tag{11}$$

In Case B, with $(1-\phi B)(1-B)^d y_t = u_t$, $|\phi| < 1$, $\phi \neq 0$,

$$\begin{aligned} RI_B(h) = & \frac{\dfrac{(1+\rho_B^{2h})\Gamma(1-2d)F(1,1+d,1-d,\phi)}{(1+\phi)\Gamma^2(1-d)} - 2\rho_B^h\gamma_y(h)}{\sum_{j=0}^{h-1}\left(\dfrac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}F(1,-j,1-d-j,\phi)\right)^2 + \beta} \\ & + \frac{2\rho_B^{2h}\beta - \sum_{j=0}^{h-1}\left(\dfrac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}F(1,-j,1-d-j,\phi)\right)^2}{\sum_{j=0}^{h-1}\left(\dfrac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)}F(1,-j,1-d-j,\phi)\right)^2 + \beta}, \end{aligned} \tag{12}$$

where

$$\gamma_y(h) = \frac{(-1)^h\Gamma(1-2d)\left[F(1,d+h,1-d+h,\phi) + F(1,d-h,1-d-h,\phi) - 1\right]}{\Gamma(1+h-d)\Gamma(1-h-d)(1-\phi^2)}.$$

The theoretical coefficients ρ *in Cases A and B are given by the expressions*

$$\rho_A = \frac{d\Gamma(1-2d)}{(1-d)\left[\Gamma(1-2d) + \beta\Gamma^2(1-d)\right]} \tag{13}$$

$$\rho_B = \frac{\left[(1+\phi^2)F(1,d,1-d,\phi) - 1\right]\Gamma(1-2d)F(1,1+d,1-d,\phi)}{\phi\left[2F(1,d,1-d,\phi) - 1\right]\left[\Gamma(1-2d)F(1,1+d,1-d,\phi) + \beta(1+\phi)\Gamma^2(1-d)\right]} \tag{14}$$

Figure 3 about here

Figure 4 about here

Figures 3 and 4 plot the relative increase in mean square forecast error for the case $\{d = 0.4, \beta = 0.1\}$, i.e., when the presence of some structure under the $\pm\frac{2}{\sqrt{T}}$ is more evident. Given the consistency results obtained by Yajima (1993) for a similar set–up, in the simulation we use the true ρ_A and ρ_B coefficients in expressions (11) and (12). From Figure 3 we clearly observe that in Case A we obtain much higher relative increases in mean square forecast errors than in the case of the misspecified white noise model, with increments of around 180% in the short–run. The same comments apply to Case B (Figure 4) with increments close to 400% in the short–run. Notice, however, that the latter case yields negligible increases in mean square forecast errors in the long–run.

5 Concluding Remarks

Due to the fact that the true data generating process of the underlying series of interest is often unknown, it is important to understand the effects of model selection. This paper studies how the choice of modeling the time series as *ARMA* process when in fact the true data generating process has long–memory affects the accuracy of the forecasts.

More specifically, herein we have examined the relative increase in mean square forecast error from fitting a weakly stationary process to the series of interest when in fact the true model is a particular member of the long memory processes, the so called *perturbed long memory* process which has the particular property of having under some circunstances all its autocorrelations below the usual $\pm\frac{2}{\sqrt{T}}$ band for all lags and for all values of the memory parameter $\alpha \in (0,1)$ and, hence, to be unidentifiable from a white noise process on the basis of the correlogram and the usual rule of thumbs in the Box–Jenkins methodology. We show that this kind of misspecification can lead to very serious errors in terms of forecasting and that corrections based on the *ARMA* family of models increase the problem, rather than provide a solution. The presence of long memory thus cannot be ignored when forecasting.

How to detect the long memory in the time series of interest? Along this paper we have claimed that the correlogram cannot be a useful diagnosis tool in the presence of long–memory time series. The same conclusion applies to the portmanteu Box–Pierce and Ljung–Box tests. In effect, Granger and Marmol (1997) prove that these test statistics are consistent against long–memory alternatives but that they could fail to detect strong dependence in finite samples and in particular in the presence of *PLM* processes.

However, Granger and Marmol (1997) show that the spectrum of the *PLM* process is unbounded at low frequencies unlike the spectrum of the *ARMA* processes. Thus, it seems that in order to distinguish between weakly stationary and long memory processes, the spectral domain can provide a more robust diagnosis tool than the autocorrelations plot.

A Proofs of Propositions

Proof of Proposition 1 First notice that $\hat{x}^{\star}_{T+h|T} = 0$ for any h, so that $e^{\star}_{T+h|T} = x_{T+h}$, and $\mathrm{var}(e^{\star}_{T+h|T}) = \mathrm{var}(x_{T+h}) = \mathrm{var}(y_{T+h}) + \mathrm{var}(\epsilon_{T+h}) = \mathrm{var}(y_t) + \sigma^2_{\epsilon}$. On the other hand, when using the correct model, the predictor is $\hat{x}_{T+h|T} = \hat{y}_{T+h|T}$ and the predictor error is $e_{T+h|T} = x_{T+h} - \hat{y}_{T+h|T} = y_{T+h} - \hat{y}_{T+h|T} + \epsilon_{T+h} = \bar{e}_{T+h|T} + \epsilon_{T+h}$. Thus $\mathrm{var}(e_{T+h|T}) = \mathrm{var}(\bar{e}_{T+h|T}) + \sigma^2_{\epsilon}$.

With respect to $\bar{e}_{T+h|T}$, from (1) we have that

$$y_t = \phi_p^{-1}(B)(1-B)^{-1}\theta_q(B)u_t = \psi(B)u_t = \sum_{j=0}^{\infty}\psi_j u_{t-j}, \tag{A.1}$$

with $\sum_{j=0}^{\infty}\psi_j^2 < \infty$ given the stationarity of y_t. Now, from (A.1) and by using elementary properties of the conditional expectation it is easy to show that

$$\mathrm{var}(\bar{e}_{T+h|T}) = \sigma_u^2 \sum_{j=0}^{h-1}\psi_j^2, \tag{A.2}$$

from which the proposition follows. □

Proof of Proposition 2 The proposition follows after some tedious algebra by using Theorem 1, Lemma 1, and Lemma 2 in Hosking (1981). □

Proof of Proposition 3 In this case $\hat{x}^{\star}_{T+h|T} = \rho^h x_T$, so that

$$\begin{aligned} e^{\star}_{T+h|T} &= x_{T+h|T} - \hat{x}^{\star}_{T+h|T} \\ &= y_{T+h} - \rho^h y_T + \epsilon_{T+h} - \rho^h \epsilon_T. \end{aligned}$$

Therefore

$$\mathrm{var}(e^{\star}_{T+h|T}) = \mathrm{var}(y_{T+h} - \rho^h y_T) + \mathrm{var}(\epsilon_{T+h} - \rho^h \epsilon_T) = \mathrm{var}(y_{T+h} - \rho^h y_T) + (1 + \rho^{2h})\sigma^2_\epsilon \tag{A.3}$$

Now, given that

$$y_{T+h} - \rho^h y_T = \sum_{j=0}^{\infty} \psi_j u_{T+h-j} - \rho^h \sum_{j=0}^{\infty} \psi_j u_{T-j}, \tag{A.4}$$

we get

$$\mathrm{var}(y_{T+h} - \rho^h y_T) = \sigma^2_u \left[\sum_{j=0}^{h-1} \psi^2_j - \sum_{j=h}^{\infty} \left(\psi_j - \rho^h \psi_{j-h}\right)^2 \right]$$

from which expression (10) follows directly. □

Proof of Proposition 4 As in Proposition 2 noting that, from stationarity

$$\mathrm{var}(y_{T+h} - \rho^h y_T) = (1 + \rho^{2h})\mathrm{var}(y_t) - 2\rho^h \mathrm{cov}(y_t, y_{t-h}). \tag{A.5}$$

Expressions (11) and (12) equally follow straightforward from the fact that coefficient ρ in (10) is equal to $\dfrac{\mathrm{corr}(y_t, y_{t-1})}{1 + \zeta}$. □

References

Beran, J. (1994), *Statistics for Long Memory Processes*, Chapman and Hall.

Breidt, F., N. Crato and P. de Lima (1998), 'The detection and estimation of long memory in stochastic volatility', *Journal of Econometrics* **83**, 325–348.

Crato, N. and H. Taylor (1996), 'Stationary persistent time series misspecified as nonstationary ARIMA models', *Statistical Papers* **37**, 215–223.

Granger, C. and F. Marmol (1997), The correlogram of a long memory process plus a simple noise, Discussion paper 97–29, Dpt. of Economics, UCSD.

Granger, C. and R. Joyeux (1980), 'An introduction to long memory time series and fractional differencing', *Journal of Time Series Analysis* **1**, 15–29.

Hampel, F., P. Rousseew, P. Rochetti and W. Stahel (1986), *Robust Statistics*, John Wiley and Sons.

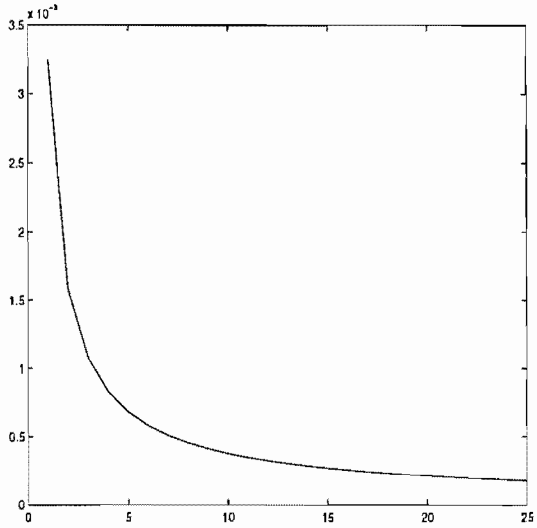
Harvey, A. C. (1993), 'Long memory in stochastic volatility', mimeo, LSE.

Hassler, U. (1994), '(Mis)specification of long–memory in seasonal time series', *Journal of Time Series Analysis* **15**, 19–30.

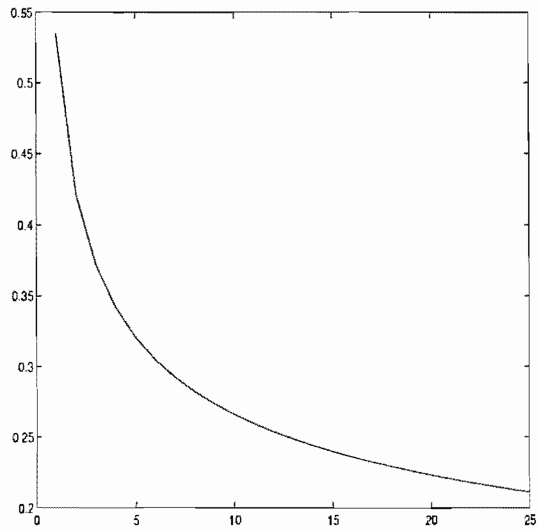
Hosking, J. (1981), 'Fractional differencing', *Biometrika* **68**, 165–176.

Ray, B. (1993), 'Modeling long–memory processes for optimal long–range prediction', *Journal of Time Series Analysis* **14**, 511–525.

Yajima, Y. (1993), Asymptotic properties of estimates in incorrect ARMA models for long–memory time series, *in* D.Brillinger, P.Caines, J.Geweke, E.Parzen, M.Rosenblatt and M.Taqqu, eds, 'New Directions in Time Series Analysis', number 46 *in* 'IMA Series', Springer–Verlag, pp. 375–382.

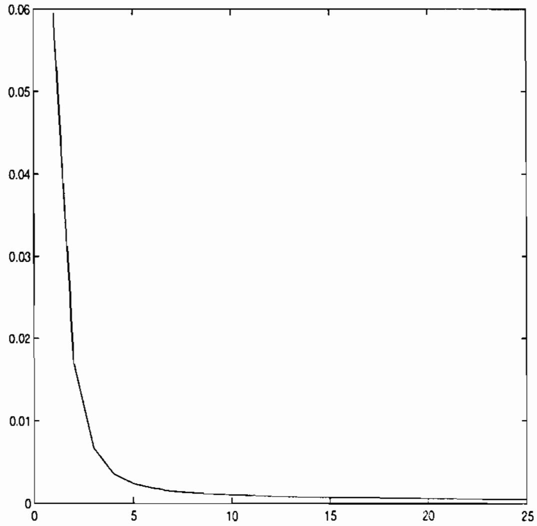


(a) $d = 0.1, \beta = 5.0$

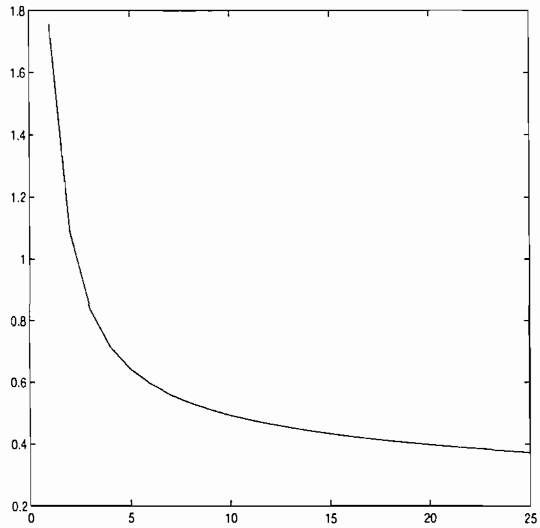


(b) $d = 0.4, \beta = 1.d0$

Fig. 1. This Figure shows the relative increase in the mean square forecast error whe the true model is given by equation (1), Case A but it is misspecified as a white noise instead.



(a) $d = 0.1, \beta = 5.0, \phi = 0.4$



(b) $d = 0.4, \beta = 1.d0, \phi = 0.4$

Fig. 2. This Figure shows the relative increase in the mean square forecast error whe the true model is given by equation (1), Case B but it is misspecified as a white noise instead.

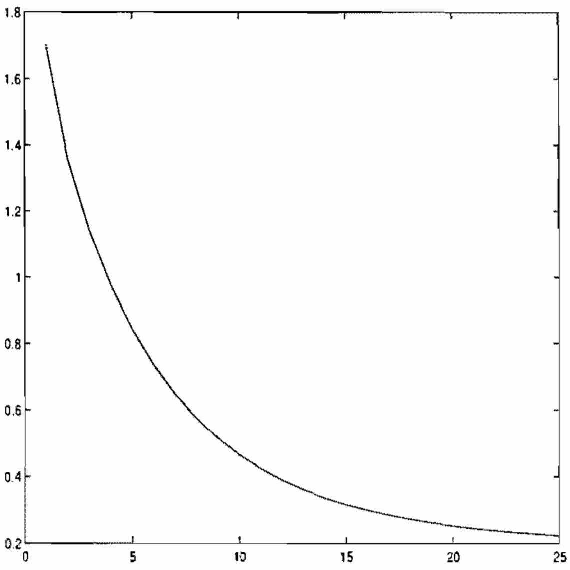


Fig. 3. This Figure shows the relative increase in the mean square forecast error whe the true model is given by equation (1), Case A, $d = 0.4, \beta = 1.0$, but it is misspecified as a $AR(1)$ process.

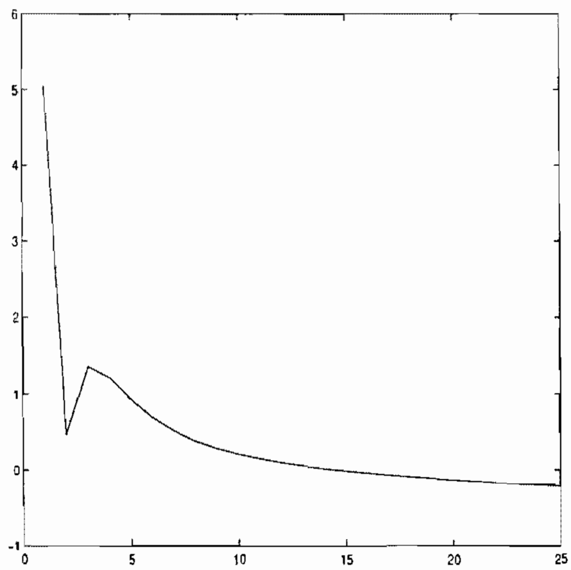


Fig. 4. This Figure shows the relative increase in the mean square forecast error whe the true model is given by equation (1), Case B, $d = 0.4, \beta = 1.0, \phi = 0.4$, but it is misspecified as a $AR(1)$ process.